East Waterway OU

Anthropogenic Background Small Working Group Meeting #1

Invitees: EPA, East Waterway Group (Port of Seattle, City of Seattle, and King County) Muckleshoot Tribe, Suquamish Tribe

November 4, 11 am - 1 pm

Agenda

1. Meeting series approach
2. Overview of technical approaches for estimating Anthropogenic Background for PCBs, dioxin/furan, and arsenic
    a. Data summing and screening
    b. Sample weighting and aggregation
    c. Summary statistic for AB
    d. Sensitivity analyses
    e. Memorandum outline
3. Initial discussion on data summing and screening
4. Work products for Meeting #2

Attachments

Small Group Meeting #1 presentation

Attendees

EPA

- Ravi Sanga
- Elly Hale
- Elizabeth Allen

USACE (on behalf of EPA)

- Bill Gardiner

Muckleshoot Tribe

- Glen St. Amant

Suquamish Tribe

Alison O'Sullivan

East Waterway Group (EWG)

- Brick Spangler (Port of Seattle)

- Jeff Stern (King County)
- Debra Williston (King County)
- Pete Rude (City of Seattle)
- Allison Crowley (City of Seattle)
- Merv Coover (ERM on behalf of the City)
- Dan Berlin (Anchor QEA on behalf of EWG)
- Greg Brunkhorst (Anchor QEA on behalf of EWG)
- Deb Chiavelli (Anchor QEA on behalf of EWG)

Meeting Notes

Dan Berlin: [Roll call] I sent around a presentation yesterday that will guide our discussion. Ravi or Elizabeth, do you have anything to add before we get started?

Ravi: nothing from me

Elizabeth: nothing profound. We had a discussion internally and some of the topics we wanted to talk about were on the agenda

Dan: Great. Let's move into the presentation.

[Slide 1] This is our first small group meeting. We want to review what we think is the right overall approach. We put this in a presentation format, but this is meant to be collaborative and we want to make sure we are on the right track.

[Slide 2, Agenda] Today we will review the meeting series approach and provide an overview of technical approaches. We'll touch on each of these topics, then we will go into the most detail on data summing and screening today. Finally, we will want to know what work products we will distribute for the next meeting.

[Slide 3, Future meetings] For today's meeting, we'll focus on data summing and screening; November 16 would be the next meeting, then we could have another meeting that week on November 20, then December 4, then December 9. Do the times look OK for folks?

Bill: for Ravi and Elizabeth, are there other schedule drivers we need to consider that would inform this meeting schedule?

Elizabeth: no issues I am aware of.

Ravi: I would like to look at this again and get back to you.

Debra: this is consistent with the discussion during the last large group meeting.

Ravi: EPA is planning to release the Proposed Plan in May. I want to review this schedule in light of that.

Elizabeth: I have not worked with you all, and depending on the group dynamics, this might be an extended schedule.

Debra: can Anchor put these dates on our calendars?

Ravi: yes, go ahead and send the invites out.

Brick: this schedule isn't set in stone and we can re-evaluate it over time.

Dan: that is correct. We would expect that we would produce a memo after the large group meeting and that the memo could be prepared like other documents with EWG preparation and EPA review. Does EPA have any other thoughts on that scheduling? If not, we have a slide on the memo outline later.

Greg: [Slide 6: reviewed slide on summing and screening] We'll go into more detail on these topics later in this presentation, so let's not open this up for discussion yet.

[Slide 7: sample weighting and aggregating] We have different flow conditions in the Green River and have different approaches for looking at the suspended solids dataset. The simplest approach is to treat all data as representative. We could also look at normalizing, as data with higher fines content has different concentration than samples with concentration with lower fines. Time-weighted and mass-weighted averages could be important to consider the flow conditions of the Green River. Does EPA have a preference?

Elizabeth: I'll give my preference. I want to hear what you think – I'm not sure that we have enough information from this dataset to do a time-weighted average with specific samples for when storm events or high flow happens a certain percentage of the time.

Greg: do you mean related to the last 2 bullets – time weighting and flow weighting?

Elizabeth: yes

Bill: I'm curious which types of sampling would be included if we are using the simple approach. Also, is there any fundamental difference between filter and suspended solids? There was quite a bit of overlap and they didn't look like 2 different populations to me. Regarding normalizing to fines, there were 3-4 samples with around 40% fines, which seemed low, but others had 60% or greater which was more reasonable. The decision to make would be whether we would normalize for fines or exclude low fines content samples. For weighting, there is a notable difference between events. Then when you couple that with the suspended solids load, there seems to be value in looking at

that. However, there might not be enough data for time-weighted approach; have not thought as much about mass-weighted approach. One of most influential factors is when we have dam release and storm, but only 1 data point for that condition. When thinking about that, I tend to agree with Elizabeth that the simple approach may be best.

Elizabeth: I don't want to throw those ideas out – part of these approaches could be part of a sensitivity analysis. I don't want to say, "let's not do it". We can discuss different ways to do it, which is always valuable. Once we have an estimate, we can see what sensitivity analysis shows.

Greg: EWG has discussed ways of dividing the data by river condition. Doing the more complex analysis could reduce the size of the dataset for some flow conditions. EWG hasn't honed in on a preferred approach.

Jeff: That is a fair statement that these are factors that we would look at. We're not necessarily saying that all of these approaches have equal merit. Time will also be an important consideration, and we probably can't admire these factors from every angle.

Debra: Does EPA have any other thoughts?

Elizabeth: I do not.

Bill: the time -weighted was the biggest one to jump out for me and also normalizing for fines and how to address differences in fines.

Ravi: Glen – anything to add?

Glen: Nothing to add at this time.

Ravi: These look good and comprehensive

Greg [Slide 8 on summary statistics for AB]: looking at AB guidance, I see reference to UCL on the mean, which is a very familiar statistic, and also background threshold values (BTVs), which can be calculated a bunch of ways in ProUCL, and a BTV can be selected from a upper threshold limit (UTL). Do you have any thoughts on this? How do we calculate a statistic, which statistic would we lean towards, and how do we compare to site performance data? EWG doesn't have a suggestion but is interested in hearing what EPA is envisioning on this topic.

Elizabeth: once we have a dataset. Calculating statistical moments is the easy part. The hard part will be comparing the condition to the site and the background data. As a general rule, we would expect that the means would be comparable. One of the ways to do that is to do a single sample comparison test to UCL; it's not a statistically valid comparison of comparing a site UCL to background UCL. Many sites would calculate a weighted mean, or a SWAC, and compare that to a

UCL, and also compare the simple mean to a UCL. Using Kaplan Meier statistic results in little variance between a UCL and a mean. I'm not entirely familiar with whether we would be doing single site comparisons or a population comparison. It depends on what intent of showing how we've achieved the cleanup goal would be done. For the LDW, they're calculating a UCL of site data to see if it meets the background number.

Greg: I'm hearing there are different ways to do it and it varies by site. UCL on the mean is a fairly common one and is our placeholder assumption.

Elizabeth: for LDW, there was a requirement that the cleanup level be measured as a UCL. I did the background calculations for Portland Harbor, and my recollection was that the value we established was a UCL on the mean of the background dataset. A BTV or UTL just tells you what is the probability that that sample is within our outside of the background dataset. We can calculate those, but those are more helpful during an RI.

Debra: EPA is using a UCL but Ecology is using 90/90UTL for background. There is a difference between how EPA and Ecology calculate NB.

Jeff: When we come back to this in future meetings, we should have some discussion on future site data and what potential interactions that would have on our choice since we haven't selected how the site would be monitored in the future yet and how it would be compared to AB. We have to wrap our heads around this issue at a certain level because future monitoring and AB affect each other.

Elizabeth: Maybe this is something that comes into RD. Ultimately, we have to have background values to develop a final ROD. As shown in this slide, there are a number of statistical values that could be calculated and put into a table in the ROD. How compliance is figured out in the future could be determined later and may not be needed now.

Jeff: I tend to agree, but we need to be aware of what limitations the selection of a background statistic puts on future monitoring. Would what is selected create significant limitations?

Elizabeth: Do you see a down-side to going down this route? Do you have a preferred statistic?

Jeff: No, I think we should list out the implications of the choices of summary statistic so we can work through that later.

Elizabeth: The more difficult task is deciding on a dataset that everyone can use.

Greg: We should also be thinking about potential future false positives and false negative scenarios

Elizabeth: Easiest way to get around that is a 2 sample test. I don't know how practical that would be here.

Bill: One topic for a future meeting would be to outline what are the considerations or limitations on the use of a specific statistic for future monitoring? What are the implications of false positives and false negatives on the ability to understand the site in the future?

Greg [Slide 9: potential sensitivity analyses.] It's important to the EWG to consider what impact inputs from laterals and post-remediation LDW bed could have on incoming concentrations. Particle size and organic carbon could affect what is settling in the EW. If we look at other methodologies for calculating AB, those could be used as part of a sensitivity analysis. In addition, other lines of evidence could provide insight into AB, such as the surface water dataset.

Elizabeth: I would prefer not to discuss other urban inputs anymore.

Elly: did we conclude we don't need a sensitivity analysis?

Debra: It was our understanding that EPA was fine with EWG looking at urban inputs as part of a sensitivity analysis.

Elly: do you have a sensitivity on how much better it could be in addition to how much worse it could be from laterals?

Debra: could you expand on that? Our thinking is that we would look at solids and chemistry weighting to provide perspective on how much they could change the incoming concentrations.

Elly: but that is not a sensitivity of the Green River

Debra: right

Elizabeth: we would want a sensitivity analysis that would affect the AB. We have already determined these other urban inputs and LDW bed are not part of the AB value. Some of the other things could be useful to decide what goes into final value. Laterals aren't useful for that. I don't see urban inputs factored into the decision-making.

Debra: we will want to have some further discussion on that.

Dan: I thought that EPA was OK with discussing the other impacts to the EW that affect long term site performance that are not included in AB. That would be a separate section of the memo that would be beyond the AB derivation sections.

Elizabeth: I can listen to any logical argument a few times. But I need an explanation of how it will be used.

Greg: [Slide 10] Slide 10 shows a preliminary outline for the memo that we will write at the end of this process.

Bill: Section 7 seems like a diversion from the primary purpose of the memo. We understand that it is a point of disagreement, and I'm not sure it belongs in this memo. I understand that this is a piece of information that the EWG wants to present, but I'm not sure how this will be used in establishing AB as a cleanup level.

Ravi: I would agree with Bill. In terms of the AB calculation, these considerations are not going to be included.

Elizabeth: whole point of calculating an AB number is because EPA believes it's achievable. If AB is not achievable, it suggests that the remedy is not performing as intended. I am not interested in hearing about technical feasibility issues; section 7 of the memo.

Greg: [Slide 12] The EW summing rules are the same as for LDW work. Regarding summing rules, non-detected qualifiers were not part of the total PCBs. For D/F non-detects, the TEFs were applied to 1/2 the DL. We've heard from Elizabeth that dioxins should be using D/F congeners not TEQ, so we would like more info on that from EPA. For PCBs Aroclors, we consider the maximum of the reporting limits when we have all non-detects.

Bill: you said the total PCB summing rules are the same as the EW RI/FS. Are they the same as the LDW?

Greg: Yes

Debra: for D/F, that is also how it was done. Each individual sample had D/F TEQ calculated. We looked at the sensitivity of using 0 vs the RL. When we did have PCB congener data, it followed the pattern discussed with NDs not included.

Elizabeth: I'm not sure on what to do with congeners vs. Aroclors – maybe we just go with a total PCB that will be the cleanup goal of the ROD. The risk-based concentration value is based on Aroclors, but we primarily have is congener data in the background dataset. I'm not sure it makes a difference in the end, unless we can show the same congeners are detected in every sample. There are a lot of substitution methods. At Bradford Island, the Army Corps came up of doing sum of PCB congeners and used Kaplan Meier to calculate a mean and multiply that mean by the number of results, which gives total sum. Interesting, but I haven't used it.

Greg: They used Kaplan-Meier statistics on each congener?

Elizabeth: no, on each sample and then calculate mean on that. That is a way of assigning values to non-detects to calculate total PCB concentration. It's an intriguing method that we can look at. For Dioxin TEQ, I don't want to calculate TEQs; we need to calculate concentrations. TEQ is meaningless as it's not in the environment.

Debra: would we calculate total D/F like total PAH?

Elizabeth: I would do it for each congener. I don't know detection frequency for each of those.

Debra: are you saying the cleanup value would not be a D/F TEQ but individual cleanup level for each congener?

Elizabeth: In Portland Harbor, we looked at what congeners were posing the most risk and calculated background for those congeners.

Jeff: for PCBs, would we be calculating a single number for total congeners and Aroclors, but for dioxins would we have 17 cleanup levels or some subset based on the highest risk? Why not just do sum of those congeners?

Elizabeth: that's a good point Jeff. Calculating a total congener D/F concentration may be best. But when you think about risk-based values, for PCBs, toxicity data is based on Aroclors and that is a single number. Summing PCB congeners probably overestimates total Aroclors PCB concentration a bit, as not all congeners are found in each Aroclor. For D/F, the highest toxicity is based on chlorines in the 2378 positions. If we do total D/F concentration, we could talk about excluding those for which we have no chlorine in the 2378 positions. It is less messy for PCBs.

Greg: is this moving backwards since we've developed the FS based on D/F TEQ?

Elizabeth: a risk-based concentration through a bioaccumulative pathway through tissue consumption is always going to be lower than background. So, going forward, we need to come up with a measure of ongoing inputs as that's what's going to be settling on the surface. We need to have like values to do these comparisons.

Debra: where do we go for next meeting?

Elizabeth: I know where we're not going, and that's a TEQ. It won't be a TEQ.

Debra: what about total PCBs based on congeners?

Elizabeth: is it a consistent congener distribution in each sample?

Debra: we know the majority tend to be the middle homolog groups, but with method 1668, there is low detection capability, we typically get detections of congeners just above DLs and can have detects and non-detects even in a sample duplicates.

Elizabeth: even though there is a low DL, non-detects have some value. We can do a Kap-sum, like what the Corps did. That way is helpful to consider non-detects without unduly skewing the result. Kaplan Meier (KM) statistics are particularly useful for this kind of analysis.

Jeff: We are seeing different PCB congener patterns showing up in different events, which we have been theorizing is different sources. Base flow has a certain pattern and have much more low molecular weight congeners (lighter ones), which might be wash off sources that come not as runoff but as what gets through the soils. Based on what we are seeing, we question whether these are source or detection issues, and how corrections could be made across different sources.

Deb: KM is meant to be applied to a single variable and estimate a distribution of non-detects, but you have to be careful applying it across different congeners. If you have different sources, may have different expectations for different congeners. You have to be careful to apply KM to a sum rather than to individual congeners.

Elizabeth: it's a variable you would have to deal with going forward and continue being used going forward for when do sampling in the future.

Merv: what is the path forward?

Elizabeth: I would like to investigate Kap-sum method for PCBs. I would like to see if it gives us a vastly different answer. We could use that for sum of detected congeners as long as that is the same method going forward.

Greg: we will do the Kap-sum and alongside what we have currently done for PCB congeners. For D/F, would we be doing departure from current approach to treat each congener as separate COC?

Elizabeth: a separate COC or calculate a total as long as we have consistent approach for calculating in the future. TEQ as an environmental number makes no sense.

Greg: why do people use TEQ in general?

Elizabeth: TEQ is used to calculate risk and have different toxic potency pegged to 2378 TCDD. You can calculate each of those separately and then apply TEF or for an individual sample, and then calculate 2378 TCDD concentration. When I do a risk assessment, I don't do TEF until calculating the exposure concentration.

Greg: if we change from a D/F TEQ to a different system, then in the future, if a concentration is too high, we might miss the actual hot spot if we're not looking at TEQs. If we do not carry TEQ forward, we might not focus cleanup on areas that would result in the most benefit to risk management at the Site.

Debra: and the RALs for the East Waterway are based on the TEQ. At some point it would be helpful to calculate TEQ to put in perspective in context of risk-based levels.

Elizabeth: TEQ should be applied on a sample by sample basis.

Debra: seems like there will be an interest in seeing what that equates to as a TEQ for considering which areas are highest risk, and that may be helpful for perspective.

Elizabeth: we could calculate each sample TEQ individually during future monitoring.

Debra: yes that would work, we would use TEQ going forward for site data but could also calculate what incoming concentration would be for perspective.

Elizabeth: yes that works.

Greg [Slide 13]: we have 7 PCB Aroclor samples that we suggest retaining, but using Kap-sum would not work with these data. Would like to present next 2 slides and get EPA reaction.

[Slide 14]: This is a log QQ plot showing the distribution of the Aroclor data. The non-detects are pretty low, but we wonder if they would still mess up summary statistics.

[Slide 15]: The 7 Aroclor data points are consistent and make sense when comparing to the congener data and considering the different event types.

Elizabeth: can you tell us your preference?

Greg: It ties into some of the other analyses. If we do congeners using Kap-sum approach, it puts the Aroclor data on the sideline. If we're keeping all the data in one bucket and not looking at flow or time weighting, then sample size is less of an issue. So from what I've heard in this meeting, the prudent approach would be using congener data and could not include Aroclor data.

Debra: but at the same time, these are real data so it's hard to ignore them, one option could be in the sensitivity analysis.

Elizabeth: summing congeners and Aroclors is never the same thing. I'm not sure that if we excluded Aroclor data that it would be a huge hit. Including it, it seems like it fits and doesn't add much noise. In Portland Harbor, as long as concentration was fairly high, using Aroclor and congener data sums, gave reasonable similar results, but at lower concentrations they were quite different. I don't think Aroclor data is a lot of data and I'm not sure it hurts to have it in and not sure we're missing anything by leaving it out.

Bill: I agree, it’s not a large dataset and does not have a large impact. Flow regime seems to matter more than whether it's Aroclors or congeners.

Debra: seems like we’re all in the middle. We may have to do a few analyses with them in and with them out.

Elizabeth: good approach, same goes for outliers if we take that data out, how much does it affect the results?

Greg: we'll keep them in and see if they make a difference.

[Slide 16]: Regarding outliers, the distribution of concentrations seem to follow lognormal distribution pattern and fairly common in the environment. We are not seeing obvious samples that should be thrown out.

Elizabeth: from a graphical example, some seemed like fairly extreme values, but not a lot. We need to consider whether the data is representative of a sample from a population. There may be some we might want to look at and consider whether it's an extreme value and consider what impact it has on the result. If it doesn't impact the result, there is no reason to evaluate further. The dataset is limited on one hand, and reasonably robust on the other. If it affects the result, how representative of the overall condition is it? Then we can decide whether we leave them in or take them out.

Debra: the trick is that higher concentrations are based on storm events and explained by higher rainfall and flow events. If we had more data to sample those storms more often, those data may not be perceived as outliers as it's the different storm conditions.

Elizabeth: they could be considered outliers as they are higher than other data, but the question is how representative should that value be when trying to come up with a value that represents the population. That one PCB value seems fairly high. If you calculate with or without that point, what's the difference? If a perceptible difference, then we'll need to talk more.

Greg: are you looking for the effect of eliminating an outlier?

Elizabeth: it's a problem having a limited dataset; how representative is the storm dataset affecting the total concentration for the year?

Greg: If that was a once a century storm and we have only 50 datapoints, then that would be less representative?

Elizabeth: and whether it's representative of the overall concentration

Deb: A once in a century may have a large impact on downstream deposition

Merv: step 1 is to see if the point is quite a bit of a departure from the other points, then do exploratory analysis whether it has a strong impact, and then make a probabilistic decision on whether to remove that point?

Elizabeth: if it's an outlier value, what does it represent? In this case we would be looking at meteorological and flow conditions. If that's the case, what happens during that condition? If it does

affect the result, and it's considered substantial, which is subjective, then we exclude it because that's what EPA guidance says to do. The smaller the sample size, the more the extreme values can affect the population. Two ways to get around it is to exclude that data point or collect more representative data.

Merv: we will need to do calculations with and without and have this conversation again?

Elizabeth: it will be an iterative discussion.

Greg: so we will pull the meteorological information on higher samples.

Bill: we can look at meteorological data. Most people agree it's a real datapoint associated with a storm event that resulted in usually high PCBs. If we're doing a UCL with and without that datapoint included, we can see what effect it has. UCLs can be sensitive to unusually high numbers, so we want to understand the level of effect. Is it a matter of including which data points to consider as a potential outlier and then running the calculation with and without?

Elizabeth: For our purposes, flow condition is more important than whether there was a storm or not. How representative of that flow condition is it of the total? Then we can do calculation.

Debra: It's more than the flow condition, it also needs to consider whether a storm is occurring, and whether the dam affects the flows. So we need to consider rainfall too.

Elizabeth: OK. The more we consider the better.

Pete: Elizabeth, when you referred to guidance, what were you referring to?

Elizabeth: 2002 background guidance, but none of it is etched in stone. It really needs to be an exercise in representativeness.

Greg [Slides 18 and 19]: we want to look at the sediment trap data and consider whether to keep samples with more fines or whether we perform a fines-normalization step.

Bill: I thought the sediment traps were out.

Greg: good to know you were thinking they were out. We were thinking they were in for individual sample comparison but out if there was a weighting approach.

Elizabeth: I was thinking they were out, but what does it do to the overall dataset if we don't include them?

Greg: OK we can present that next time. We will also bring information related to summing, looking at Corps approach for handling non-detects for PCB congeners.

Elizabeth: I will send a memo on this process

Greg: For dioxins, we will be looking at individual congeners. For outliers we will be evaluating the representativeness of the data.

Debra: we will be running things with and without outliers and then using flow and rainfall data to understand.

Greg: I thought I heard flow and rainfall first and then calculated UCLs with and without second.

Elizabeth: more information is always helpful; we don't have to use it all.

Bill: for D/F, are we talking about looking at congeners for background and then calculating background TEQ, or just calculating separate AB value for each congener?

Greg: I don't know we settled on this. It would be nice to keep a TEQ background value with applications for managing site risk. We will look at individual congeners to not pigeon-hole anything; we will look at congeners and then also calculate TEQ

Jeff: regardless of which way it ends up for background values, PP/ROD would need to turn it into a TEQ value to compare to risk values to characterize projected risk of cleanup that can be achieved

Greg: We will look at the data in both ways.

Dan: Do we want to calculate statistics or are we not there yet?

Ravi: I would like meeting minutes and identify work products for what EWG is responsible for. Will you provide meeting minutes?

Dan: yes, do you want us to provide additional raw materials on sharepoint? Bill, you have been looking at the data, do you need anything else from us?

Bill: I don't need anything else, what you're providing to the group is fine.

Elizabeth: I think this conversation is really helpful. If we don't have calculations completed next meeting, it probably won't affect schedule. After we decide what dataset is, statistics will be easier.

Jeff: we will need to look at how to prioritize these analyses

Elizabeth: Unresolved things include weighting with fines and weighting for flow conditions. Particularly for flow conditions, we have not discussed how we might do that. It would be best to have more discussions before doing calculations.

Debra: let's put that on the schedule for next meeting.

Ravi: we accomplished a lot, but have a long way to go.

Dan: we will schedule meetings and send out meeting minutes.